

# Scaling Laws for Generative Mixed-Modal Language Models

Year: 2023 | Citations: 135 | Authors: Armen Aghajanyan, L. Yu, Alexis Conneau, Wei-Ning Hsu, Karen Hambardzumyan

---

## Abstract

Generative language models define distributions over sequences of tokens that can represent essentially any combination of data modalities (e.g., any permutation of image tokens from VQ-VAEs, speech tokens from HuBERT, BPE tokens for language or code, and so on). To better understand the scaling properties of such mixed-modal models, we conducted over 250 experiments using seven different modalities and model sizes ranging from 8 million to 30 billion, trained on 5-100 billion tokens. We report new mixed-modal scaling laws that unify the contributions of individual modalities and the interactions between them. Specifically, we explicitly model the optimal synergy and competition due to data and model size as an additive term to previous uni-modal scaling laws. We also find four empirical phenomena observed during the training, such as emergent coordinate-ascent style training that naturally alternates between modalities, guidelines for selecting critical hyper-parameters, and connections between mixed-modal competition and training stability. Finally, we test our scaling law by training a 30B speech-text model, which significantly outperforms the corresponding unimodal models. Overall, our research provides valuable insights into the design and training of mixed-modal generative models, an important new class of unified models that have unique distributional properties.